Draft for Comments

Chapter: Multimedia Authoring

Text Book on

Fundamentals of Multimedia Computing

Authors:
Gerald Friedland
and
Ramesh Jain

Draft for Comments

Chapter: Multimedia Authoring

Introduction

Multimedia data is used to communicate different aspects and perspectives of information related to an event or an object. The concept of document has been used over centuries as a device or mecahanism to communicte information. Since the technology to store and distribute information has been evolving and changing, the nature and concept of document has been evolving to take advantage of the new media technology. At one time, one considered a document to be in the form of physical embodiment such as a book and mostly contained text as the source of information. This popular notion of a book as a document has gone through changes over the last few decades and has now transformed the notion of document to be a (virtual) container of information and experiences using digital data in multiple data formats. In this modern reincarnation of document, it is not limited to one medium, but can use different media as needed to communicate the information most effectively to a recipient.

In this chapter we discuss types of different multimedia documents. We present concepts and techniques behind many established as well as emerging systems for preparing multimedia documents. Our emphasis is in presenting concepts and how they can be applied rather than presenting details of a particular product. Creation of multimedia documents has been a very active area and there are many popular products. One can learn how to use those products from books and manuals on those specific products. In our discussion, we will not cover specific details of any product.

What is a Document?

The most common widely used document is a book. Gutenberg's invention of moveable printing press popularized the concept of a book by facilitating creation and distribution of books. Even today, when somebody talks about a document, most people think about a book. However, to the generation of people growing up with Internet and the WWW, a book will evoke a different image. They will consider a book to be collection of text and images, presented by a person to make it coherent and complete, but frozen at a particular time. Since a book was printed on a paper with substantial efforts and cost involved, it could not be easily modified or updated. Each subsequent edition was once again carefully thought about and prepared to make it complete and remain current and complete for a foreseeable future. A book was divided in multiple chapters. Each chapter addressed a particular topic again trying to be complete on that topic. In most cases a book organization could be represented as a tree structure, as shown in Fig. 1. This tree structure was mapped into pages. One may consider pages as necessary

structure imposed by physical requirements. On one hand, the text and images in a book have to be readable and hence of some minimum size. On the other hand the whole book should be such that a normal person should be able to handle it. This can be easily accomplished by designing a book as a series of attached pages so that one could flip them in a sequence in which the material in the book is presented.

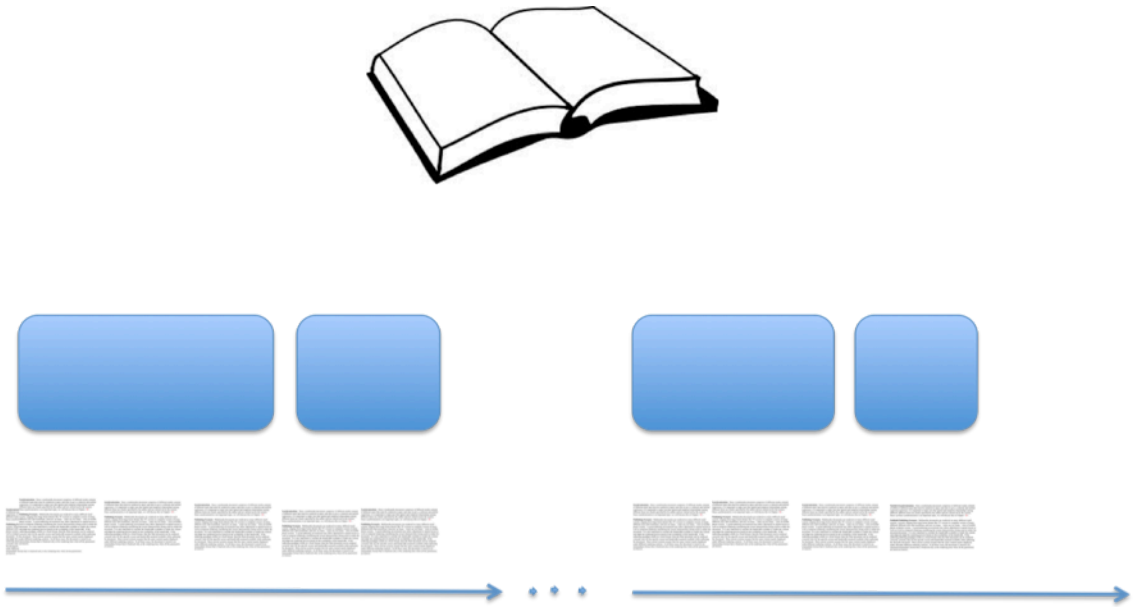


Figure 1: Organization of a book. Each page really takes an interval of the text stream and converts it into visual representation that has to be spatial – a page. Then pages are assembled into chapters, which in turn are assembled into a book.

A very important thing to note in a book is that the text is now limited to pages. As we considered earlier, text is a symbolic representation of speech. And speech is a temporal signal. Thus, one may consider that a book is created by considering a timeline and dividing into appropriate intervals, which are converted into pages. Each page is a folded version of the timeline.

The concept of representation of temporal information visually on a page became dominant, if not the only, mode of creating documents due to the technology that was available. Different techniques were used to emphasize important part of the text, such as bold face, larger fonts, and different colors. Different spatial layouts were used to emphasize different aspects of text.

Arrival of audio and video technology changed the book metaphor for documents. Audio and video allow rendering of time varying audio and video signals in their natural audio and video form. The limitation of having to artificially represent time as visual pages is no longer constraining the types of documents that could be rendered. Now a document can contain time-varying signals in time varying form.

Another major transformation in documents is the linear nature of documents. Paper based physical representation forced linear structure on books. A book could be theoretically read in any order, but authors prepared a book assuming that usually it will be read linearly, from beginning to the end. Some people read books in somewhat random order, but most books, particularly fiction, are read in a linear order. Arrival of hyperlinks in electronic documents and then introduction of hyperlinked pages in the Web changed this notion. As we will see, now a document could be read in an order that a reader finds appropriate rather than what the author intended it to be. More importantly, now a document is no more a compact and closed physical artifact, but a dynamic organically growing artifact that could present multimedia in all its forms. And we have only seen the beginning of how future documents will be.

Evolving Nature of Documents

Most documents may be considered as a composition of many *content segments* (CS). A CS is a component that has been either authored or captured and can be considered an independent unit of media that could be combined with other segments. It is like an atomic segment that could be combined with other units to build increasingly complex documents. A CS could be a text document, a photo, a video segment, an audio segment, or any other similar data that represents a particular media. Each CS has associated meta-data that provides essential context related to its interpretation, rendering, utilization, and authorship. What is stored in meta data is dependent of the media and application domain. Some meta-data elements that have become *de facto* standard across different media are size, name, date, and place of author or device acquiring the data, and coding method to convert the media to bits. We discussed EXIF for photos earlier in Chapter <context>. One may want to look at the meta data related to text files or other data on any system to get a good feel of how meta data looks. In Figure 2, we show some content segments and elements of meta data associated with those.

File Type	Common meta data associated with the file
Text Document	Name, Author, Length, Date-Created, Date-Last-Modified,Type, …
Photo	Name, Capture-Time, Compression-scheme, EXIF, …
Audio	
Video	Name, Creation time, Compression, Length, Type, …

Figure 2: Content segments of different media and meta data commonly associated with those: a. Text document, b. Photo, c. Audio, and d. video

Almost always, the meta data about a CS is not rendered when the segment is presented, but it is always used in deciding the rendering method. Also, whenever one wants to use a particular CS, meta data is used to determine its relevance and how it could be combined for a potential new document. It is important to understand that without meta data, a segment may become unusable.

Given several relevant CS for producing a document, one may combine them in many different ways. Different combinations may result in different documents. To understand different ways to combine these, let us consider 10 different segments shown in Figure 3. In Figure 4, we show three possible combinations in which a document may use it. The first composition approach used in Figure 4a combines them in a fixed linear document that is rendered, using conventions of English text, in left to right sequence in time. In Figure 4b, these documents are composed by the author as sub-documents that could then be used as new CS that are then rendered linearly. It is possible to combine different components in many different ways. In Figure 4c, we show linking of documents so a user can go from one composite document to other if she so wishes by breaking the strict sequence that is followed in 4a and 4b.

One may consider evolution of documents along three important dimensions:

Type of Media: Until recently, most documents were mostly text documents and were commonly available in printed form on paper. Occasional photos or figures were included to enhance understanding of concepts or details that were considered too complex for text.

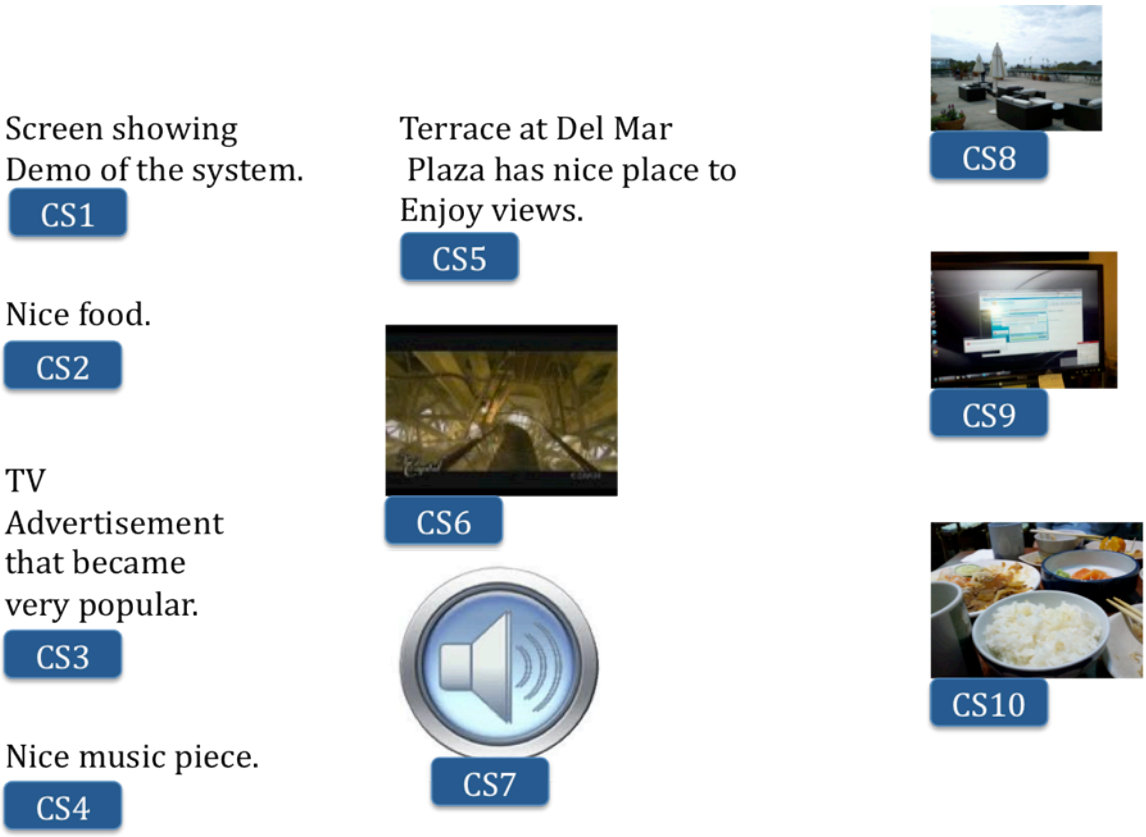
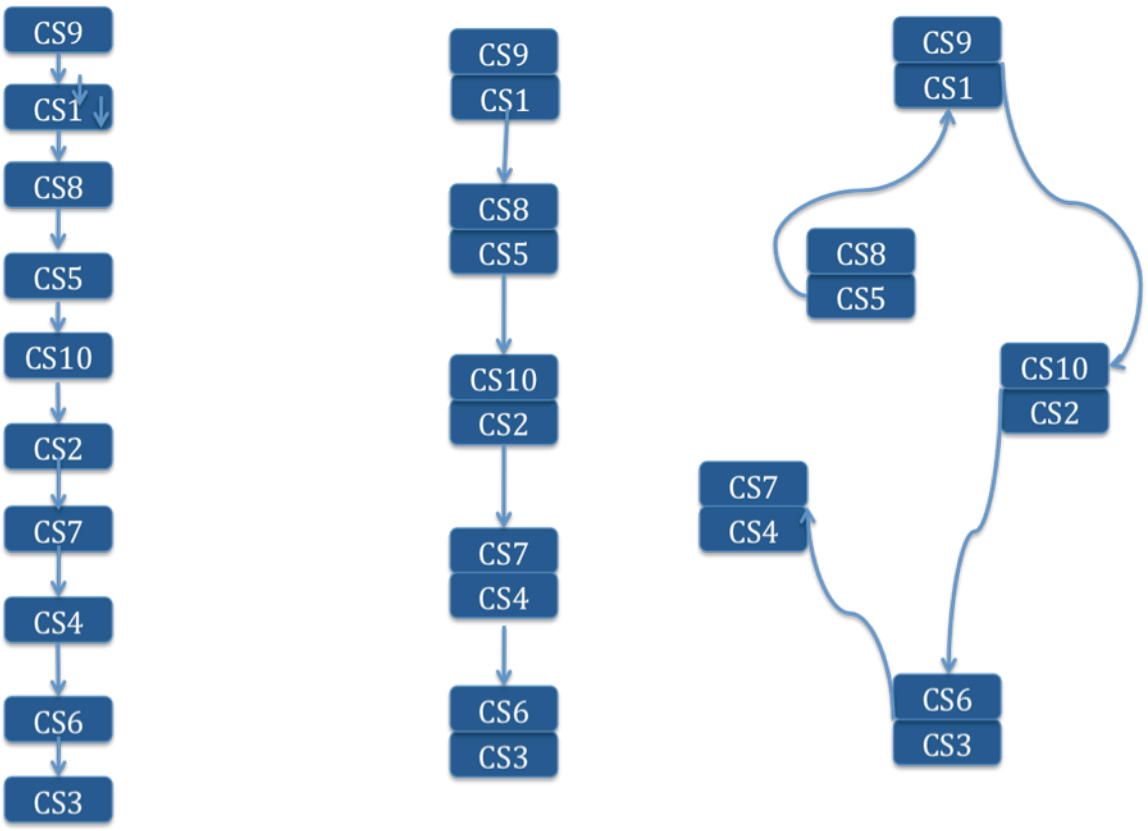


Figure 3: 10 different media segments. Each of them is independent unit and is considered a atomic unit.



4a **4b** **4c**

Figure 4: Three different combinations of the segments of Figure 3. In 4a, the traditional linear segment is presented; 4b shows some atomic documents combined to form composite elements and then used in the document; and 4c shows how a user may navigate from one unit to the other as she may wish.

In the last few years, due to emergence of new media technology, the nature of documents has gone through a complete metamorphosis. Text is and will remain an important component of documents, but now documents use different media as the author deems suitable for communicating the information and experiences in the best possible manner. Different media can be combined in space and time in appropriate manner by the author to communicate his ideas in the most compelling manner. Moreover, same CS could be used as many times in a document or by as many documents as need it. This is now possible because unlike in old days, a CS is in electronic form and hence could be copied and used effortlessly or linked for rendering it without copying it. This ability has resulted in revolutionary changes in creating new documents and provided new powerful approaches for expressing ideas.

Non-Linear Flow: Due to the nature of physical documents, text was designed to flow linearly. This was strongly influenced by the temporal nature of narrative structures natural to text-based story telling. With the advent of electronic media and ability to create links, the limitation of linearity can be easily overcome. At first thought people

accustomed to linear structures find nonlinearity confusing and unnatural. However, the fact that one can compose independent CS and then can link these to form multiple linear structures using different links, is making this approach very popular. As we will see in the following, this provides very flexible approach to authoring documents that may be customized for different types of audiences.

Dynamic Documents: Older documents required significant efforts to create and then were distributed using static medium such as paper. This resulted in a significant latency between an event and sharing information and experiences of the event. During early days, the latency was really large. Newspapers were invented for reducing this latency between an event and its report for important events. Television brought live events but resource limitations kept this limited to only important events. The Web brought blogs, micro-blogs, and now real time automatic updates for sharing live event information and experiences. There is an increasing trend to compile a document related to an event as it is unfolding.

In addition to such event reports being dynamically unfolding with events, they can also be designed to suit information relevance and needs of a person. This is resulting in creation of personalized dynamic doscuments.

Stages in Document Creation

Every document is created using a three step process. These three stages do have some overlaps, but are distinct enough to consider them separately.

Data acquisition and organization: When a person decides to create a document, she starts thinking about the information, experiences, and message that the document will communicate to a user of the document. This involves thinking about the relevant events and related information that must be used in the document. In some cases this information is already available to the user, while in other cases, this must be acquired. In most cases, the information available is significantly more than that could be used in the document due to the size limitations of the document. The size limitations of the document are due to attention span of the user; more than the physical requirement which in earlier times were more dominant.

A major change brought on by technology in the last few years has been the increasing availability of meta data that helps in organizing and using the data. Meta data is available for text files, as well as photos, videos, and all other data that is created or collected. In most cases, meta data is stored with the data in the same file. This could be used for organization as well as for using the data.

Selection: An author[1] of a document usually collects lots of material in preparation for conveying the message through the document. It is common to for an author to collect significantly large volume of material in anticipation of its use in the final document. All this material must be organized so that it is available to support the author in selection of all pertinent material. Many meta-data management tools have been developed to help potential authors to organize and select such material. **<Give this information in further reading>**

The author selects the material from the content segments in the database considering the message that needs to be conveyed. The factors considered in selection of the segment are: relevance of the segment to the message, length of the segment, media of the segment, and how this segment could be combined with other segments.

This step is usually an iterative process. Once initial material is selected, the author must consider which material should be included in the final document. The author must consider the type of media available to convey the same information and experience and which one will be the best in the given context. Another important factor to be considered in the iterative process is the length of the document as well as the effectiveness of the information and media used to convey that. The output of the process is a set of segments to be included in the final document.

Editing: Editing is the process of taking an existing document and modifying it for its use in a given context or simply for refining it. Editing is media dependent. Many powerful tools have been developed for editing documents of specific media type. Here we briefly discuss some common operations used in commonly used professional tools.

Text Editing: Many tools have been developed for editing text documents. Commonly used operations involved in editing a text document are

- Insert,
- Delete,
- Format to change the layout, and
- Emphasize using different styles, sizes, and colors.

The first two operations are obvious for changing the text. Formatting is used to provide structure to the text and includes breaking the text into sections or paragraphs, adding footnotes or references, creating special textboxes, and similar things to provide clear visual separation on a page. The final operation of emphasizing is to clearly display relative importance of certain text segments by using bold, italics, underline, larger font, different font styles, or different colors. Human beings use different intonations and inflexions in oral communication. Since text is a static representation of oral

---

[1] We will use the term 'author' for the person who prepares a document. In some cases, like for video, usually the term 'producer' is more common. But we will use author for all kind of documents.

communication, these emphasis tools are used to capture some of the characteristics of oral communication.

Photo Editing: A photo is a flat static representation of visual information. Most photos are captured using a photographic device, but there are other mechanisms such as computer graphics or human painting or sketching used for creation of photos. Some of the common operations used in photoediting are:

- Selection of important objects,
- Addition of Objects,
- Deletion of Objects,
- Enhancement or restoration of visual characteristics, and
- Changing visual characteristics in parts of a photo.

In photographs, the most important aspect is to clearly mark pixels in an image that may represent a particular object. This operation is significantly more difficult than it appears. Many tools such as magic wand and cropping are provided to facilitate this operation. Enhancement and restoration are aspects that are normally used to compensate for some artifacts introduced due to imperfections during the photo capture operation. Visual characteristics are changed in parts of photo to make visual appearance more appealing. Finally, addition and deletion of objects are fundamentally to change the content of a photo. A photo represents state of the real world captured at a particular time. By inserting or deleting an object, an editor is changing the state of the world as depicted by the photo[2].

Audio Editing: While many older audio editing tools try to simulate a tape recorder, modern editing operations on audio are usually based on a visual representation of the amplitude space (going from left to right in time). Audio can be

- cut out,
- copied,
- pasted, and
- filtered.

The problem with most visual representations of audio recordings is that they are not intuitive, e.g. the user must listen in very often as the amplitude space representation does not indicate the final acoustic experience good enough. Several tracks are usually visualized above each other. Speech editors therefore often show a spectrogram (see Chapter XXX) of the speech signal, allowing a more intuitive representation for experts. Midi editors allow the editing of notes, which makes it easier for musicians. They work with Midi editors like a text processing tool.

---

[2] Before photo editing tools became common, some editing was done in dark room. Before digital tools arrived, a photo was considered a strong evidence of what the state of the real world was at the time photo was taken. Digital photo editing tools allowed manipulation of photos and eliminated photos as an evidence of the real world.

Video Editing: Video is different from the above media in that it combines all of the above and adds some new dimensions. It has spatial dimension and characteristics of photos, but represents rapidly varying sequence of photos thus bringing in temporal dimensions. It is also combination of not only a photo sequence but also of audio that is either captured with the video or is added to the photo sequence. Moreover, one could either overlay text in some parts of video or even use text exclusively as video segments. Video editing tools usually contain:

- Photo editing tools
- Adding a video segment at a particular time say Ti
- Deleting a video segment from time T1 to T2.
- Add a text box at specified location from time Ts to Te
- Add an audio channel from time Ti
- Add an imagebox a specified location from time Ts to Te
- Add another video, usually of a much smaller size, at a specified location from time Ts to Te
- Add specific transition between segment Sk and Sk+1.

As can be easily seen, video editing tools utilize results of editing of all other media and must provide spatial and temporal composition operations to combine different media to provide a coherent media. In Figure 5 we show a video authoring/editing environment that contains photos, video, and audio components that are combined using timelines.

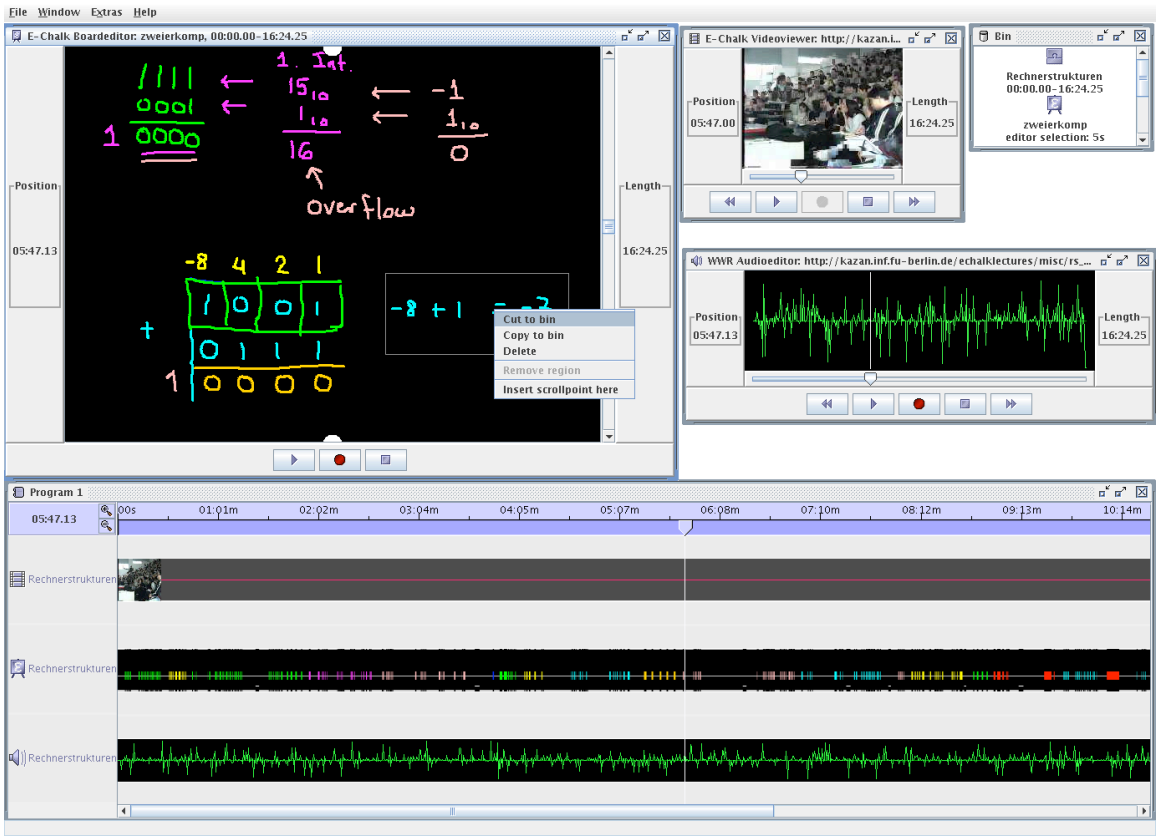


Figure 5: A video authoring/editing environment uses a timeline for showing how different components can be organized.

Emerging Multimedia Editing Tools: In a way multimedia tools are extension and collection of individual medium editing tools. Since in most of the current multimedia systems also a screen is used for rendering, spatial layout is manipulated similar to text and photos. Audio and video bring in temporal elements. This means that tools must be provided to manage time. Multimedia editing tools are similar to video editing. The major difference in multimedia and video editing tools, however, is that video editing tools consider the photo sequence as the driving medium and other media play a supporting role. In multimedia, video is considered at the same level as any other media. In fact, a good multimedia authoring environment considers that all media are equally important and one must use a specific medium to convey or emphasize important information or experience that is most relevant for communication. A multimedia authoring environment must consider elements discussed in the following section.

Basic Elements of a Multimedia Authoring Environment

Since a multimedia document utilizes all media to make a document that combines appropriate medium to communicate the message in the most compelling manner, it must provide facilities to author each individual medium and to combine them effectively and efficiently. Moreover, to facilitate interactivity of the user with the document, an authoring environment must also consider mechanisms for user interactivity at the time of authoring. Based on the emerging changes in the nature of documents, one must consider different factors in designing multimedia authoring environments. Two very important fundamental aspects are related to spatial and temporal composition of different media assets. As we see below, one must pay careful attention to layout as well as synchronization issues.

Characteristics of Media Assets: Different media assets have different spatial, temporal, and other informational attributes that play important role in the combined documents in terms of designing their layout and synchronization. In many cases these characteristics are stored as meta-data along with the data corresponding to the media. In some dynamically created content, this meta-data as well as the data becomes available only at the rendering time. An authoring environment should account for this.

Spatial Layout: Most multimedia documents in current systems are rendered on a screen. The screen has fixed spatial dimensions, such as 640 X 480 pixels or 1920 X 1200 pixels. An author decides which media item should be displayed in which area of the screen and what should be its resolution. In some cases an item must be scaled up or down to fit the selected window size. Thus, with each media item, there is an associated spatial window where it should appear. In Figure 6, we show the screen and multiple

windows. Each window size must be specified using either a rectangle in absolute locations or in terms of a corner and its size. For each window, one must also specify the type of content and the source from where the content must be displayed.

Figure 6: A spatial layout showing three different windows. Each windows location and size must be specified. Windows A1 and A2 are within A; and C1 and C2 are within C.

Temporal Layout: Since multimedia content can be displayed as video on the screen, an authoring environment should specify different content that will be used to constitute this video. The earliest authoring tools in this area started appearing in video editing systems. Multimedia authoring systems extend them to include more sources of data and provide more flexibility and control in using and combining the data. An example of temporal layout is shown in Figure –<temp-layout>. This layout essentially provides tools to specify the time interval for the appearance, transitions, and disappearance of each content item on the screen. The location of the content on the screen could also be specified.

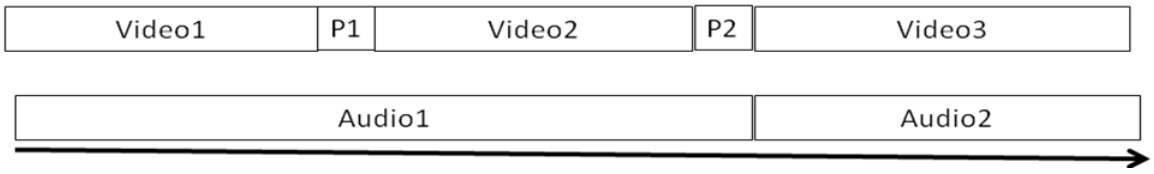


Figure 7: The timeline representation of each content item is shown on the timeline. For each item the start time and the duration must be specified.

Synchronization: A multimedia document comprises of different media contents of different types that must be rendered in space and time to give a coherent and unified experience. It is important to make sure that spatial and temporal relationships among different items are clearly specified by the author and are carried out by the system. Since synchronization is an important topic, we will discuss this in Chapter <*X*>. Most multimedia authoring environments provide basic tools to specify which media elements should be synchronized.

Publishing Formats: Multimedia documents are rendered on many different sized screens, ranging from large home theater like TV screens to computer screens of many different sizes and resolutions, and now on many – some say too many – sizes of mobile phone screens. A good authoring environment may allow adjustment in spatial layout as well as temporal rendering considering the screen characteristics being used to render the document. It is also important to consider the bandwidth available to render the content and adapt the rendering process based on the availability of the bandwidth. If the authoring paradigm results in a fixed format, then the final document can be rendered correctly only for the specific screen and bandwidth assumed available while authoring the document. Most current systems assume that the same content maybe displayed under different rendering contexts. In all these cases, the document is stored in an intermediate format that is finalized only at the rendering time when all the parameters are known.

Representation of a Multimedia Document

As may be obvious from the above discussion, the structure of a multimedia document is relatively complex. A text only document has fairly linear structure comprising of chapters, sections, and subsections. With modern hyper-linking capabilities, nonlinearity has been introduced in otherwise linear text documents. Now a user may play a role in defining the rendering of these documents also, as discussed earlier in this chapter. Due to flexibility in organizing spatially and temporally and use of multiple types of media, the nature of the multimedia documents becomes relatively more complex to understand.

Many different models have been used to represent multimedia documents. In this section, we discuss two of these models that cover many requirements of multimedia authoring environments and have gained popularity.

Structure-based Representation: Common structure-based representation uses a tree structure in which the root is a complete document and the leaf nodes are individual media elements or a pointer to those. Intermediate nodes are 'sub-documents' comprising of combinations of individual media elements. For each intermediate node, the composition rules and spatial and temporal layouts maybe explicitly specified. Figure 8 shows structure of a multimedia document that contains multiple text, photos, audio, and video segments. One content element could be used multiple times if desired.

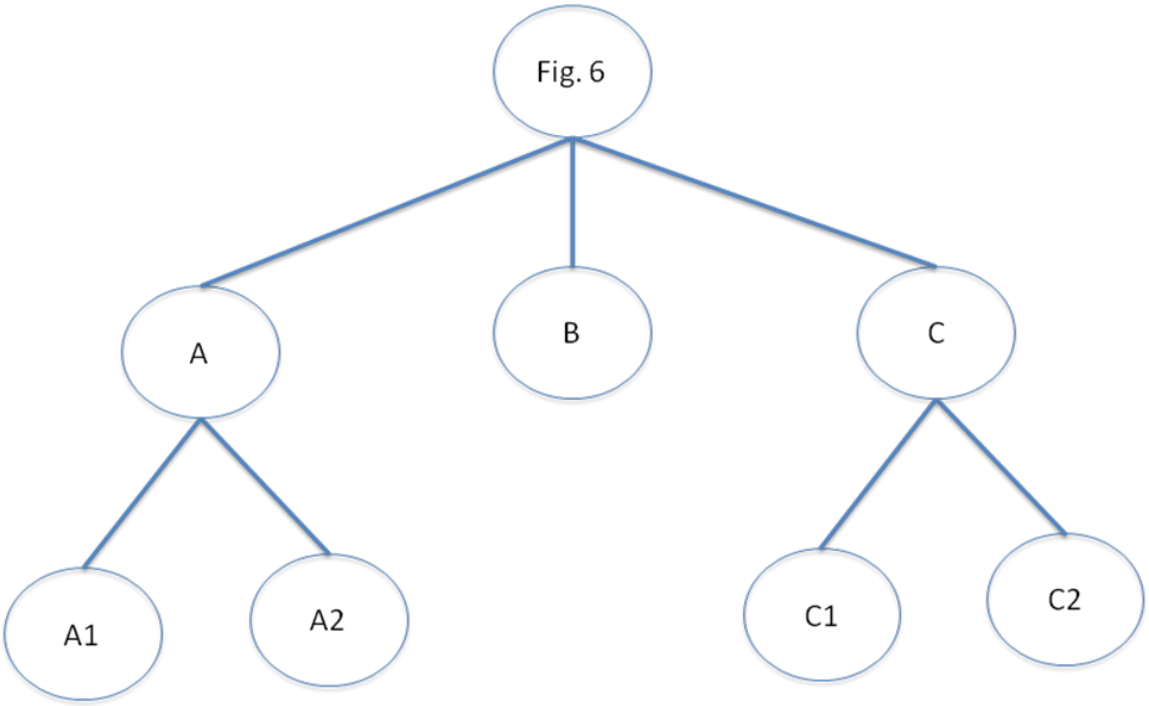


Figure 8: The tree structure shown represents a complete multimedia document that uses several media components. This structure represents the window shown in Figure 6.

Time-Based Representation: Time based representations evolved from video editing. In these representations one considers that a multimedia document is organized around a timeline. Different media elements are represented as different tracks synchronized with the master timeline. Each track specifies which content element will appear during which time interval. One may also specify the relative spatial position of each media element. This representation makes the relative appearance and disappearance of different media elements explicit and easy to represent and understand.

Current Authoring Environments

Many multimedia authoring environments have been developed in the last two decades. One of the biggest drivers for developing these environments has been the Web. Many other systems were also defined for general multimedia authoring environments, such as MPEG4 and SMIL. Rapid convergence is taking place in devices, communication, and computing. It appears that the Web environment may become the unifying environment. In the following we briefly discuss some key concepts and trends among emerging authoring environments.

HyperText Markup Language (HTML) was defined to be the first publishing language of the Web and has remained the main language for preparing documents so they can be published on different platforms. Like any other markup language, HTML uses tags to specify how an element on a page should be published. The language syntax defines how to specify tags for different actions to be performed. These tags are in pairs like <T1> and </T1>, where the first tag declares the beginning of T1 and the second tag is the *end tag* closes it. Most of the information in text, tables, and images is between the tags. The tags are used by a browser to interpret the intent of the author in displaying the content of the page or the document. In the early days of the Web, most of the documents usually contained only text. Tags in those days usually specified presentation related operations on text. HTML1.0 was a key component of the launch of the Web and was predominantly concerned with presentation of the text on a page.

As the nature of documents changed to more multimedia, subsequent versions of HTML provided specifications for inclusion of multimedia content. These specifications had richer tags for layout of media items. Another challenge faced by browsers in the presentation of multimedia content was use of proprietary technology for playing video. The latest version of HTML, HTML5, has introduced specific features to author multimedia content as easy as text. In particular, it now has four specific constructs: <video>, <audio>, Canavas, and SVG. These features make inclusion of multimedia content in a document much easier than earlier.

Further Reading

An excellent history of development of diffeent media and the impact on society is presented in 'Cognitive Surplus' by Clay Shirky. James Gleick's book, The Information: a History, a Theory, a Flood, is an excellent source for the changing nature of information and how it has affected our society.

Photoshop was a major force in converting photos from a visual record to an authoring environment. Photos, often called images, used to be a record that could be processed to enhance them and to recover some information. By providing simple tools to edit them, Thomas Knoll's system, developed when he was a doctoral student supervised by Ramesh Jain at the University of Michigan, changed the way photos were viewed. From a a record it became a creative environment for expressing visual thoughts. The impact of Photoshop on multimedia authoring is not only in photos, but also in video production. On the lighter side, Photoshop destroyed what used to be considered an irrefutable evidence – a photo of an event – and has now resulting in creation of multimedia forensics as a field.

An important multimedia authoring project that contributed many important ideas and resulted in development of a complete multimedia environment was Synchronized Multimedia Integration Language (SMIL). This environment developed over several years and made available to community was one of the first authoring environment to consider all aspects of multimedia authoring and make it compatible to emerging concepts and tools from the Web community.

MPEG4 was the first effort to consider video as composition of objects and events both for compression as well as for providing interactive environment dynamic visual environments. Efforts started in creating multiple perspective and immersive video in the 20th century, but due to technology limitations remained in the conceptual stages. With advances in technology, it is expected that these techniques will advance rapidly and will result in powerful immersive telepresence systems.

Finally, one is seeing emrgence of new media as a new communication mechanism for knwledge. Emerging social media systems rely on combination of multiple media to communicate and share experiences, unlike the dominant medium of text that started with Gutenberg's moveable printing press and has remained dominant so far.

References

‘Cognitive Surplus’ by Clay Shirky

The Vanishing Line Between Books And Internet

Hugh McGuire http://bit.ly/crDZok

James Gleick, The Information: A History, a Theory, a Flood, Pantheon books, 2010.